KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Trade-offs in operating room planning for electives and emergencies: a review

Van Riet C, Demeulemeester E.

KBI_1426

Trade-offs in operating room planning for electives and emergencies: A review

Carla Van Riet*, Erik Demeulemeester

KU Leuven, Faculty of Business and Economics

Department of Decision Sciences and Information Management

Naamsestraat 69, B-3000 Leuven Belgium

Carla.VanRiet@kuleuven.be

Erik.Demeulemeester@kuleuven.be

Tel: +32-16-32.69.72

Tel: +32-16-37.90.72

*Corresponding author

Trade-offs in operating room planning for electives and emergencies: A review

The planning of the operating rooms (ORs) is a difficult process due to the different stakeholders involved. The real complexity, however, results from various sources of variability entering the processes. These uncertain processes cannot be ignored since they greatly influence the trade-offs between the hospital costs and the patient waiting times. As a result, a need for policies guiding the OR manager in handling the trade-offs arises. Therefore, researchers have investigated different possibilities to incorporate non-elective patients in the schedule with the goal of maximizing both patient- and hospital-related measures. This paper reviews the literature on OR planning where both elective and non-elective patient categories are involved. It shows the various policies, the differences and similarities in the research settings and the resulting outcomes, whether they are beneficial or not. We find that the dedicated and the flexible policy are mostly pursued, but the setting and the assumptions of the reviewed papers vary widely. Decisions on both operational policies as well as on capacity are required to assure timely access and efficiency, which are the two main drivers for the problem at hand. Furthermore, the policy choice impacts the number of schedule disruptions and the OR utilization. However, results on the overtime and the patient waiting time are partly contradicting. The review shows that some policies have already received considerable attention, but the question of which policies are most appropriate is not yet fully answered. Neither has the full spectrum of policies been explored yet. Consequently, this topic provides several areas for future research, which are outlined throughout the paper.

1. Introduction: Sources of variability

Ideally, the healthcare sector would be able to deliver the highest quality of care at the lowest cost by providing the right resources at the right time to the right patient. Unfortunately, the life of healthcare providers (and patients) is made difficult due to all kinds of variability. Examples of events inducing variability in the complete surgical process include [13,30,62]:

- Late arrivals of patients
- Late arrivals of medical records
- Late or early arrival of medical staff
- Delay in support services
- Inaccurate reservation of resources
- Setup, clean up or change over time variability
- Acute onset of abnormal medical conditions (e.g., infections)

- Surgery duration variability
- Uncertainty in duration (length of stay) of all upstream and downstream activities
- Arrival of emergency patients

Many of these aspects determine whether or not the operating rooms (ORs) will run out of time [45]. Variability is important since the OR schedule influences several other departments in the hospital, such as the intensive care units (ICU), the wards, the laboratories and the Emergency Department (ED). As such, the daily variability in the elective OR caseload is the main cause of ED diversion to other hospitals [57] or of the variability in the downstream resources [33].

Litvak et al. [58] introduced the terms 'natural' variability and 'artificial' variability, which were later adopted by other authors. The former consists of variability due to the different types of diseases, each with a varying degree of illness (clinical variability), due to the unpredictable arrival of patients (flow variability) and due to the differences in the professional abilities (professional variability). This natural variability drives up the cost of care, can hardly be avoided and thus must be optimally managed. 'Artificial' variability is both non-random and non-predictable [57]. Here, many causes can be possible including patient preferences or practices of the provider. One example is the day-to-day variation in the elective scheduled caseload, which is introduced into the system by the scheduling process. This variation covers more than 80% of the admission and occupancy variation from the OR [58]. Artificial variability disorganizes the system since an efficient organization, where supply is nicely matched with demand, is made impossible. Haraden and Resar [33] even report that the effect of artificial variability caused by personal preferences and beliefs of the surgeons far exceeds the natural variability.

From the stochastic aspects listed before, the literature focuses mainly on the three last ones: surgery duration uncertainty, uncertainty in length of stay (LOS) (or bed availability) and arrival of emergency patients. First, surgery duration variability can be countered by having good estimates (e.g., [95]) or by for instance planning the expected duration increased by an amount of slack, in order to avoid overtime with a certain probability (e.g., [32,39]). Huschka et al. [39] show that this considerably affects patient waiting time without greatly affecting patient throughput or OR utilization. Secondly, the LOS variability can be partly reduced by having a master surgery schedule (MSS) that takes the LOS into account (e.g., [8]). As this paper focuses on the OR, the LOS will be further excluded from the analysis of performance measures. Finally, the arrival variability can be tackled in different ways, which is the main focus of this review.

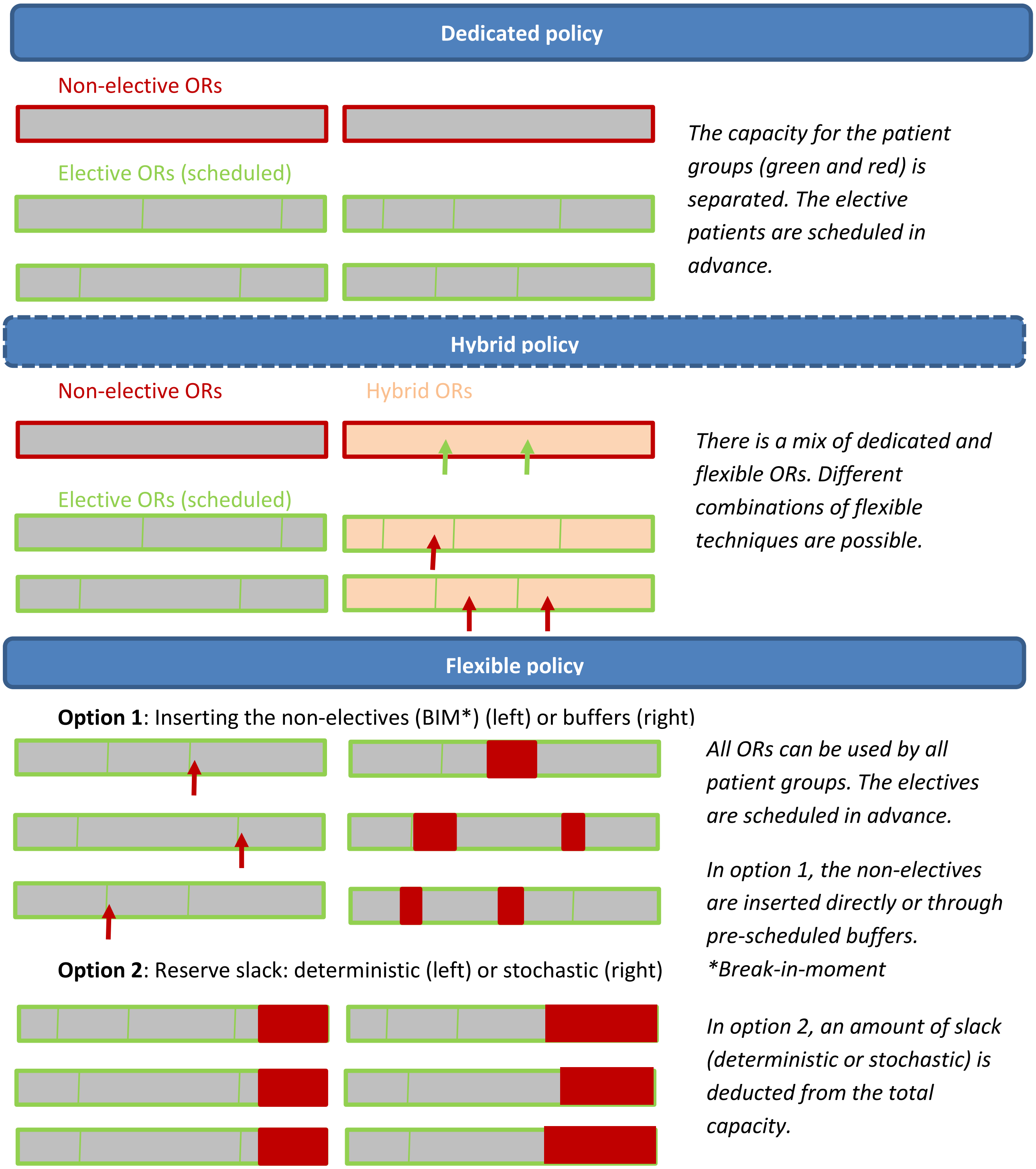
Emergencies are a key cause of variability and they need to be taken into account in order to guarantee sufficient capacity. Unfortunately, the literature on how to include non-elective patients is scarce. Cardoen et al. [13] confirm that only limited research is being done on non-elective patient 'scheduling'. Since non-electives are intrinsically difficult to plan, most literature on operating room planning only reports scheduling practices for electives [30]. Moreover, only 29% of he reviewed papers by Guerriero and Guido [30] consider stochastic aspects. Most of these papers use tailored heuristics to overcome the computational challenges (e.g., primal-dual heuristic, column generation-based heuristic). In addition, most researchers use the variability from the data in their model, but focus less on the reduction of the variability in the process [13]. Clearly, these uncertain processes cause a need for policies to guide the decisions of the OR managers on how to manage the planning of the ORs.

2. Tackling the trade-offs

The difficulty in scheduling patients in general is the trade-off between cost and efficiency on the one hand and the quality of medical care and patient's preferences on the other hand. The unpredictable nature of emergencies and the fact that they should be served on short notice creates an extra trade-off between allocating operating theater resources to non-elective patients or to elective patients. More specifically, since electives can be scheduled in advance, hospitals pursue a high efficiency level reflected in high utilization rates, acceptable patient waiting times and short turnover times. However, for emergencies, responsiveness or quick access is required. Ferrand et al. [26] discuss this trade-off in healthcare and other domains.

To deal with this well-known trade-off, three policies for handling emergencies are pursued: the flexible, the dedicated and the hybrid policy. In the flexible policy, there is no separate OR reserved for non-electives and several rules and strategies are used in order to manage the access for the two patient categories. In the dedicated policy, one or more ORs are dedicated to a specific patient type in order to separate the flows of the patient categories. The hybrid policy is a combination of both in which for instance some capacity is reserved for non-electives, but other ORs are also accessible by non-electives. The different policies, illustrated in Figure 1, are further discussed in more detail.

Figure 1: Illustration of the three different policies for handling non-electives



3. Organization of the review

In this review, we focus on the OR literature that directly impacts or considers non-elective surgeries. It includes both research on tactical allocations (capacity) as well as on operational patient scheduling. Although all policies cover a mix of capacity decisions and scheduling decisions and therefore a clear distinction is often not feasible, option 1 in Figure 1, i.e., inserting non-electives or buffers, is considered closer to the operational level while option 2, i.e., reserving slack, is considered to be tactical research. At the tactical level, the focus is on having fast access to care for the non-elective patients while minimizing the disruptions for elective patients. Operational studies focus more on the best way to schedule both patient groups. When a paper discusses more than one policy, it is classified under the policy that receives the most attention in the paper.

Mainly papers that use operations research techniques (mathematical modelling, simulation) and data-analysis are discussed. Other techniques such as workflow management and business process re-engineering (e.g., [6]) are not included. Non-electives in the up- and downstream resources such as the ED and the ICU (e.g., [59]) are not included in the classifications. Papers that assume dedicated capacity for non-electives, but do not further take this patient category into account in the model (e.g., [67,89]) or papers only considering one patient category are also excluded in the classifications.

In the patient scheduling literature, there is a lack of consistent designation of patient categories. The following terms are all used to describe patients who cannot be scheduled well in advance: emergent, urgent, add-on, work-in and semi-urgent. We will use the term non-electives throughout this review to address this patient group.

The rest of the review is structured as follows. Section 4 provides an overview of the characteristics of the building blocks that are used in the literature to cope with the scheduling of non-electives. Section 5 discusses the literature on the flexible policy and sections 6 and 7 treat its dedicated and hybrid counterpart respectively. Conclusions and future research opportunities are discussed in the final section.

4. Building blocks characteristics

When comparing papers, the researched setting and the corresponding assumptions are important. This section describes these building blocks of the literature on non-electives in the ORs. First, the type of analysis and the applied solution techniques are summarized. The next two subsections discuss the policies in relation to respectively the scope of the research combined with the size of the case hospital, and the time window of the dataset combined with the decision level of the research. The fourth subsection looks at the modeling characteristics. Since categorization and prioritization are important aspects for non-elective 'scheduling', the fifth subsection clarifies these aspects. Finally, the ratio of non-electives to elective patients is reviewed.

In a hospital, non-elective patients follow a specific flow. They usually enter the OR complex from a variety of locations (e.g., the ED, the inpatient wards, the ICU or an external hospital). Once the patient arrives, an OR and a surgical team must be booked and the current elective schedule might need to be adapted. After surgery, non-electives follow a pathway similar to the one of electives.

4.1. Type of analysis

The overview of applied solution techniques and the corresponding type of analysis, summarized in Table 1, shows the popularity of mathematical programming, simulation modelling and statistical analysis. In the medical journals, data-analysis and simulation are often used to tackle OR problems of a specific case hospital. In the case hospital often the situation before and after the implementation of a specific OR policy is examined. In the papers applying operations research methods, simulation, mathematical programming, heuristics and queueing analysis are mostly used. The following paragraphs discuss the popular methods in more detail.

Table 1: Type of analysis and solution methods

Type of analysis	
Scenario analysis/modelling	[1,10,23–25,32–34,52,63,65,68,70,72,75,80,82,86–88,90,93]
Compare before-after	[7,9,21,34,55,60,74,78,80]
Optimization	
Exact	[1,23,28,46,49,52,70,72,73,84,94]
Heuristic	[1,19,22,23,32,37,47–50,52,65,96]
Complexity	[23,32,47–50]

Solution method	
Simulation	
Discrete event simulation	[1,19,23–25,52,65,68,70,72,75,82,90,93,94]
Monte Carlo simulation	[10,32,47–49,65]
Data-analysis	[7,9,21,34,55,60,74,78,88]
Analytical analysis	
Queuing theory	[33,34,63,70]
Markov theory	[86,87,96]
Other	[36,48,49]
Constructive heuristic	[19,23,32,37,47,48,50,52]
Improvement heuristic	[47,50]
Meta-heuristic	[22,23,32,48,52]
Mathematical programming	
Goal programming	[1]
Column generation	[47,50]
Mixed integer programming	[23,47–49,65,72,73,94]
Linear programming	[46]
Branch and Bound	[84]
Dynamic programming	[28]
Unclear	[80]

Note. The classification schema is based on the one of Cardoen et al. [13]. Since [23] and [93] are follow-up articles on [52], the last one will not be classified separately in the remainder of the paper.

Discrete event simulation (DES) is a common approach for studying complex environments and for taking uncertainty into account. Not surprisingly, it has been increasingly popular as a tool to tackle problems in the operating theatre, which is reflected in the increasing number of scientific contributions using this method for research on ORs. As shown in Table 1, simulation is also commonly used to model problems specific to non-electives. It is used for modeling the (non-) elective patient flow, the resource allocation, the patient scheduling and capacity decisions. This result fits in the findings of Brailsford et al. [12], who show that simulation is prominent in planning and resource utilization problems.

Several reviews on simulation modelling in general (also including Monte Carlo simulation, system dynamics etc.) in healthcare are published (e.g., [41,42,64]). Moreover, Augusto et al. [4] provide a framework or meta-model for simulations of healthcare systems and Jahangirian et al. [40] discuss the lessons learned from the commerce and defense sector that might be applicable for simulation in healthcare.

Whether or not to use simulation depends on the research objectives. The goal to incorporate the stochastic elements of the ORs as precisely as possible often motivates the use of simulation models. Simulation can provide answers to particular questions of the case hospital and is suited for scenario analysis. On the contrary, Brailsford [11] discusses the barriers to implementation of the simulation models in healthcare and Jahangirian et al. [40] show that the sector is lagging behind on that aspect compared to the other two discussed sectors. Both papers mention the resistance to (organizational) change and the data capture as causes for failed projects.

Mathematical programming and mixed integer programming in specific is a popular optimization technique. It is mainly used in the reviewed papers to create optimal operational surgical schedules. As a main benefit, it can yield solutions which are optimal, given the input.

Queueing analysis has been applied to various healthcare problems, especially the appointment scheduling problem. There are several good reviews of queueing theory applied in healthcare [27,29]. Queueing analysis is not only suitable for assessing the operational performance of a system, but is also used for determining capacity requirements, such as determining the required number of ORs for certain patient categories. Both applications are found in the non-elective literature. Moreover, McManus et al. [63] show that queueing theory is suited for modeling critical care resources. Also trade-offs between OR utilization and overtime or cancellations are modelled using queuing theory (e.g., [96]).

Furthermore, Markov models (both Markov chains and Markov decision problems (MDP)) can be used to model or optimize healthcare problems. A problem is Markovian if the future state depends only on the current state and not on states preceding the current state (i.e., the feature of memorylessness). Markov chains are used to model patient flow, can be used as statistical models of a system and are popular in for instance ED literature. Tancrez et al. [87] use Markov theory to guide decisions on the size and the allocation of capacity to (non)-electives.

An MDP is essentially a sequential decision model and adds decisions and rewards to the Markov Chains. The MDP has been widely studied by mathematicians in the last two decades, but so far it has hardly been used to cover significant applications [38]. Zonderland et al. [96] use an MDP to determine at the beginning of a particular week how many slots should be planned in that week for the so-called two-week semi-urgencies, which are urgencies that must be served within two weeks.

4.2. Scope and policy focus

Most of the reviewed papers look into a problem of a specific case hospital. Therefore, it is important to look at the setting of the research and the applied policy simultaneously, as shown in Table 2. In general, both the dedicated and the flexible policy have received attention in the literature. The dispersed classification of the papers in Table 2 partly provides a reason for the contradicting results on the policies in the non-elective literature, especially when it comes to the question of whether or not to dedicate ORs to non-electives.

Table 2: The applied policy (focus) and the scope of the research

	Scope				
Policy	**All departments**			**Specialized**	**No clear data**
OR size	<10 ORs	10-15 ORs	>15 ORs		
Dedicated	[87]	[55,70,93]	[24,34,75,80]	[9,60,72,74,78]	[7,86]
Flexible					
Option 1	[19,23,73][a]	[23]	[19,23,32]		[84]
Option 2	[22,87]	[65,93]	[24,36,37]	[1,46,96]	[28,47–50,86]

[a] Two hospitals are researched in [19].

More than one third of the papers, classified in Table 2, research a setting with one or only a few departments. This setting or scope is different from the one including all departments. Clearly, different departments or pathologies have different characteristics in terms of arrival patterns, duration patterns (both mean and variance) and allocated capacity and staff. Isolating a pathology can therefore greatly influence the results and even more importantly, limit the results to the characteristics of this pathology.

Furthermore, the size of the OR complex of the hospital greatly influences the results as confirmed by van Essen et al. [23] and is therefore shown in Table 2. Indeed, a size of less than ten ORs provides fewer possibilities to reserve a full OR for non-electives compared to a hospital with double this amount. The majority of the papers that consider all departments research an OR complex of reasonably large size. Note that although the number of beds is regularly used to depict the size of the hospital, Table 2 shows the number of ORs since this is more relevant in this context.

Within the dedicated policy even more than half of the papers that could be classified according to the chosen classification, research a specialized department of the hospital. In the majority of the cases, the orthopedic department is discussed. Since the specialized departments might already be selected because of their favorable characteristics, results have to be interpreted with caution.

Finally, option one of the flexible policy provides an opportunity for future research, since it has received limited attention so far. Especially research on inserting variable-sized breaks at several spots in the schedule is scarce. Moreover, only one paper incorporates data of hospitals with different OR sizes.

4.3. Time window of the data and the decision level

As mentioned before, the two main policies for handling non-electives are studied at both the operational decision level as well as at the capacity decision level, as reflected in Table 3. Some papers discuss both levels and focus on the operational scheduling policy on the one hand and the number of dedicated ORs (DORs) or capacity on the other hand.

For the dedicated policy, most research focuses on the tactical decision level (e.g., how many ORs need to be dedicated). Operational policies to manage the non-electives in the DORs are lacking and form an area for future research.

Once the decision level is known, the presented data must be examined carefully. Unfortunately, they are often not fully disclosed, which makes it hard to check the quality of the dataset. Nevertheless, the time-window covered by the data can be an important aspect to keep in mind when interpreting the results. For instance, the proportion of papers that use data of less than one year is surprisingly high for the tactical decision level, as shown in Table 3. Of the papers covering one year or more, the majority works with data of one year.

Table 3: Time window of data and decision level per policy

	Operational		***Capacity***	
Time-window data	< year	>= year	< year	>= year
Dedicated	-	[72]	[24,34,60,87]	[7,9,55,70,72,74,75,78,80,93]
Flexible	[19,73]	[1,23,32]	[24,87,96]	[1,36,37,46,93]

Note.[65]: unclear time-window.

4.4. Modelling characteristics: Surgery duration and patient arrivals

Variability in durations and arrivals are main causes for scheduling difficulties. Therefore, the corresponding assumptions determine the range of analytical possibilities and limit how well the model resembles reality. Table 4 provides an overview of the distributions or techniques that are used to model the durations and the arrivals of elective and non-elective surgeries. Authors that are not mentioned in the table did not include information about the distributions.

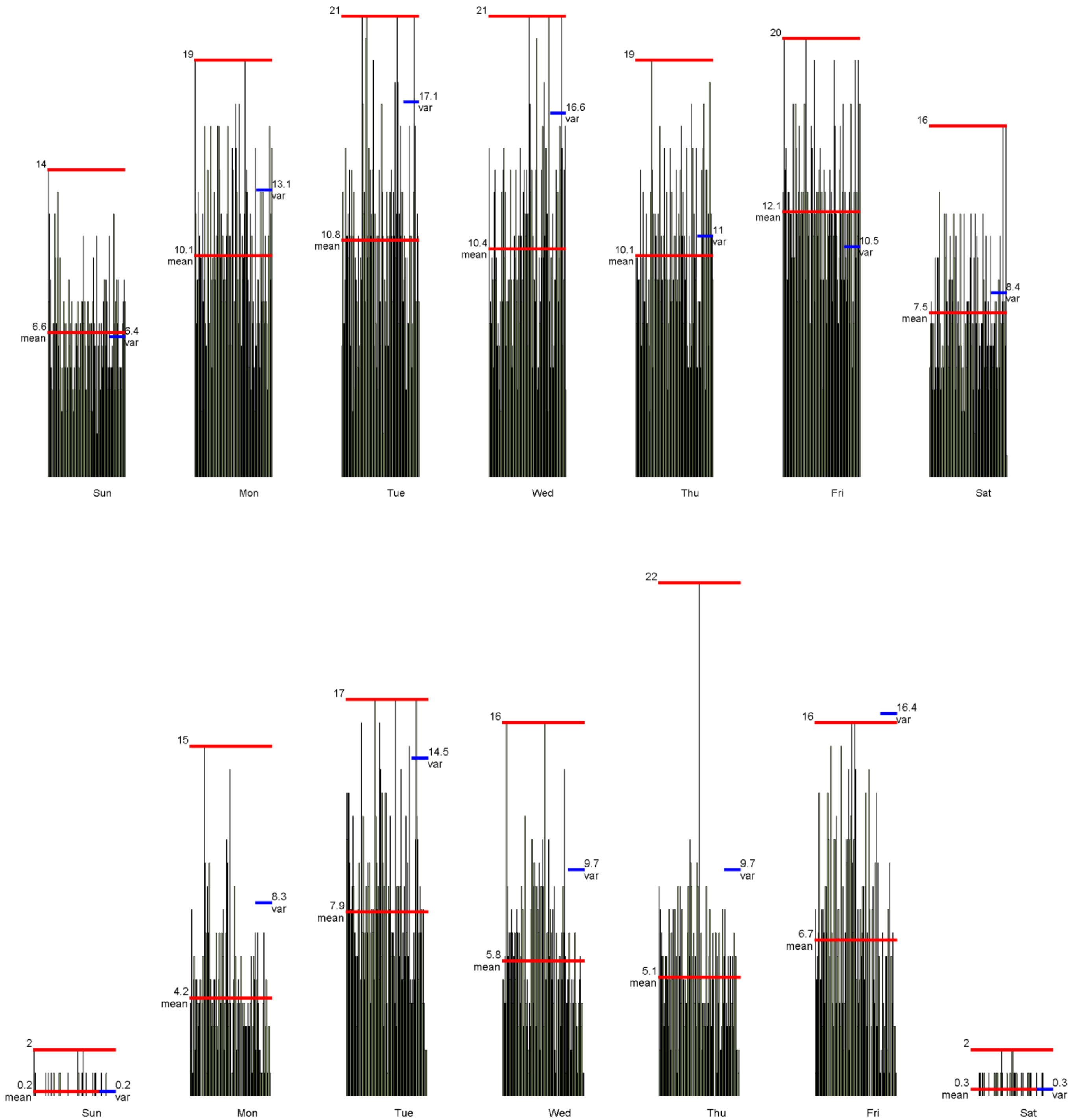
Table 4: Modelling assumptions on durations and arrivals for both electives and non-elective**s**

	Duration		Arrival	
Elective	• Lognormal	[23–25,47,72,82,84,93]	• Poisson	[1,72,94]
	• Deterministic	[1,46,49,50,73]	• Deterministic	[1,23–25,93]
	Historical mean	[22,23,72,82]		
	• Empirical distribution	[65,75]		
	• Uniform	[48,49]		
	• Normal	[28,32,36,37,65]		
	• Exponential	[86,87]		
Non-elective	• Lognormal	[19,23–25,47,72,82,84,93]	• Poisson	[1,10,23–25,70,72,86,87,93,94,96,97]
	• Deterministic	[1,48,96]	• Deterministic	[1]
	Historical mean	[72,82]		
	• Empirical distribution	[10,75,94]		
	• Erlang	[70]		
	• Normal	[28,36,48]		
	• Exponential	[48–50,86,87]		

Note. Papers based on data-analysis (see Table 1) report the realized durations and arrivals and are therefore not included, papers on the hybrid policy are.

With regards to arrival times, two approaches can be discerned. Firstly, the expected number of arrivals per weekday, also called the arrival rate, can be used to model the patient arrival process. This arrival process is generally modelled as a Poisson process. This Poisson distribution implies that the mean arrival rate equals the variance. However, this assumption is far from true in certain settings. As an example, the data of a large Belgian hospital with 22 ORs [77] is shown in Figure 2. For a discipline like Urology, the variance clearly deviates from the mean. For the non-electives, shown at the top of Figure 2, the situation looks slightly better although the Poisson assumption does not hold for each weekday.

Figure 2: The number of arrivals for a given weekday per week (104 weeks for 2012-2013) for non-electives (top graph) and Urology (bottom graph) as well as the corresponding mean, maximum and variance [77]



A second approach to model the arrivals uses the inter-arrival times. The inter-arrival times are especially popular in simulation models. Following the results from the previous paragraph, the exponential distribution is often assumed. As an example, the inter-arrival times in the Belgian Hospital are shown in Figure 3 [77]. The graphs in Figure 3 show the maximum likelihood estimator (MLE) and the MLE with the top 5%, considered as outliers, removed. The exponential distribution (blue line) does not fit the data very well for most disciplines as explained before.

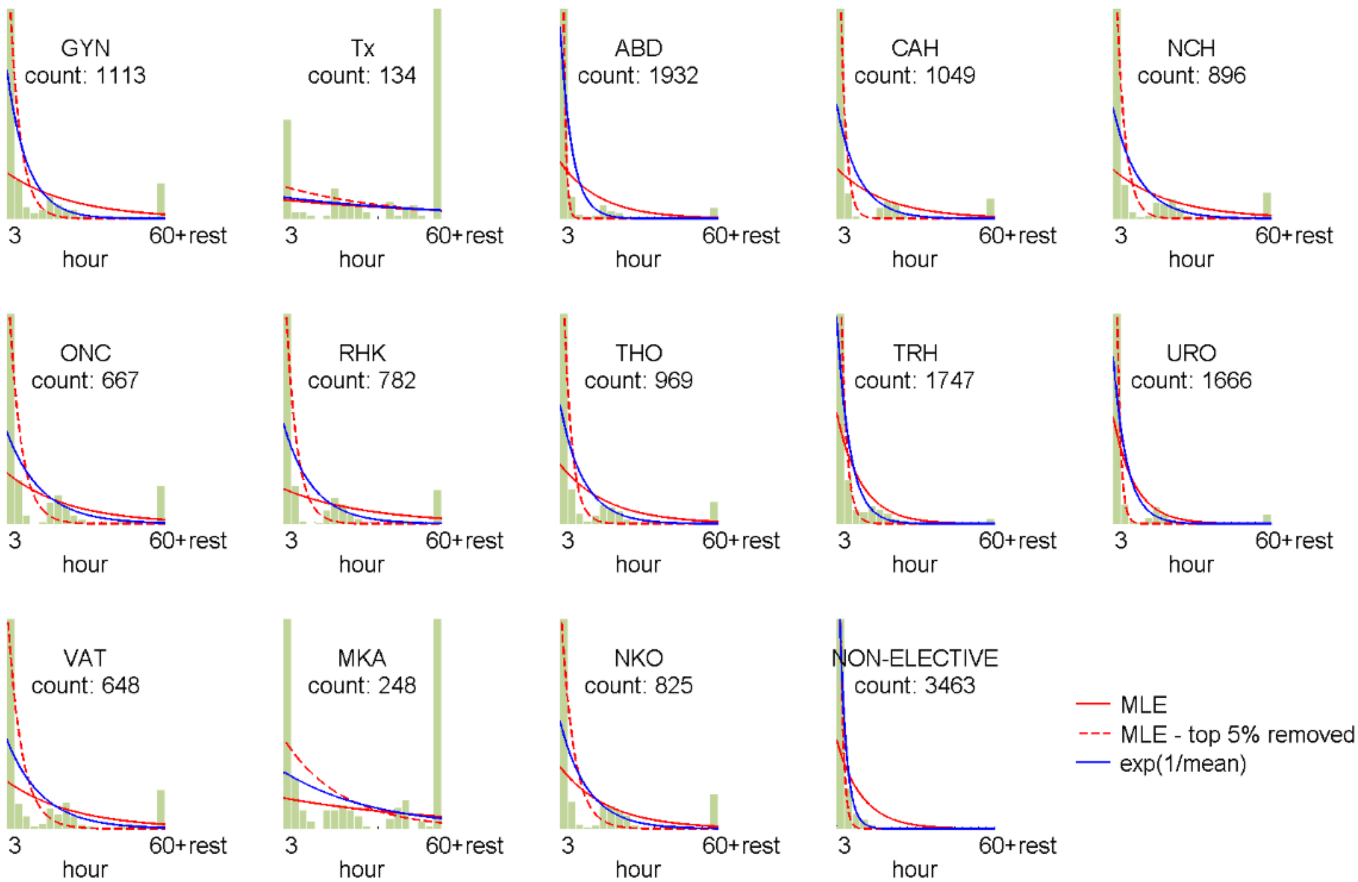
The use of the exponential distribution brings along several benefits which (partly) explains its popularity. As such, the parameter for the distribution can be easily calculated by taking the inverse of the mean. Moreover, several theoretical queuing results are available for systems with exponential inter-arrival times.

The lognormal distribution often results in the best fit for the hospital data on surgery duration, as shown in Table 4, which is also confirmed by Strum et al. [83]. Moreover, deterministic surgery times, for instance the historical mean for a specific surgery type, are also common. Clearly, this last assumption might lead to unrealistic results due to the implicit assumption of no uncertainty.

The estimation of the surgery duration remains a topic that requires attention in practice. Lebowitz [53] even reports inaccurate estimations as the most important cause of the lack of OR punctuality. However, Dexter et al. [17] show that optimally choosing the operating day for the elective patients is more important in order to best fill the allocated hours compared to eliminating the errors in the estimation of the surgery duration. A detailed analysis of surgery duration estimation is out of the scope of this paper.

Although the assumptions on arrival and duration characteristics are crucial for interpreting the results, several papers fail to specifically mention them. Reporting all assumptions clearly should receive special attention in future research. Time-varying arrivals could also be considered more often in future research.

Figure 3: Inter-arrival times of patients per specialty (with 3 hour bins) [77]



4.5. Categorization and prioritization

If we refer to different patient categories, the question of how to categorize and prioritize them raises naturally. Litvak et al. [58] argue to separate the different sources of variability by classifying the patients into homogeneous subgroups. Since a trade-off exists between separation and flow variability, one cannot endlessly divide the patients into smaller subgroups. Therefore, finding a good basis for categorization is crucial. Examples are disease type (e.g., Cardiology, Orthopedic), attending surgeon (related to professional variability), disease severity (e.g., complex or simple), urgency (elective or non-elective patient) or resource type (outpatient or inpatient).

A wide range of categorization definitions is used. Table 5 provides the different categories and the accompanying service target. Unfortunately, the same category is sometimes used for denoting patients with varying service targets. Although a classification system can be important to decrease the waiting time for non-electives [56], these patients are often categorized under a general term like 'emergent' or 'urgent'. Especially for the add-on cases a common definition is lacking, since this term is used both as a collective term as well as to categorize a specific type of patient.

Moreover, often no (interval) target time is reported. Nevertheless, van Oostrum et al. [68] provide an example of the efficacy of introducing safety or time intervals during the night shift. The intervals positively influence the number of non-electives treated within their deadline. Moreover, the target time can provide hospitals a way to measure the category-specific waiting time performance.

The vague and inconsistent categorization makes it more difficult to compare or benchmark papers in this field. Furthermore, a review on the advantages or implications of different classification systems is an opportunity for future research.

Table 5: Categorization of the non-electives

	Target	**Reference**
Trauma	Now	[10]
Emergent	<30min	[24]
	<1h, <2h	[70,73]
	<6h	[9]
	<24h	[72,87]
Urgent	<4h	[70]
	[6h, 24h]	[9,73]
Semi-urgent	<8h	[70]
	<1/2w	[96]
Add-on/non-urgent	<24h	[9,73,75][a]
	<1w	[80]
Work-in	[24h–1w]	[75]
Other		
Priority levels (P)	P1-P3: <1h, <4h, <12h;	[34]
	P1-P3: <8h, <8-24h, <24-48h	[55]
	P1-P3: <6h, <24h, <5d	[51,78]
	P1-P5: <8d, <30d, <60d, <180d, <360d	[88]
	P1-P5: <45min, 2h, 4h, 8h, 24h	[56,70,75,80]

Note. w= weeks; h = hour; min = minutes; d = days.
[a] [73]: Cases to fill-up capacity for the next day.

Categorization can be done based on several aspects as mentioned earlier. Many papers (and thus case hospitals) use the medical priority as main categorization basis. A recent example of categorization based on medical priority is one where patients are assigned a time interval in which their surgery is medically advised, called the due time [91]. An example of the corresponding categorization is shown in Table 6. The due time concept also turns up in Canadian, British and Italian research. This categorization recognizes that scheduling lower-priority patients later in the time horizon provides additional flexibility to schedule higher-priority patients [69]. Similarly, Sandbaek et al. [78] show that introducing a new classification system resulted in a system where 90% of the patients could be planned ahead.

Table 6: Categorization as applied in the university hospital of Leuven [77]

	Category	Meaning	Interval
Non-elective	1	Now!	[0, 0]
	2	Up to 6h	[0, 5h]
	3	Today	[0, 23h]
Elective	4	1 week	[1d, 7d]
	5	1 – 2 weeks	[8d, 14d]
	6	2 – 4 weeks	[15d, 28d]
	7	4 – 8 weeks	[29d, 56d]
	8	8– 16 weeks	[57d, 112d]

Conversely, questions can be raised about the quality of the common medical categorization, as reflected in Table 5. In practice, most of the non-electives can actually be treated within 24 hours. Typically only 15% of the non-elective patients must be served within six hours of admission [51]. Similarly, Heng and Wright [34] argue that only 9% of the patients are of the highest urgency category (i.e., requiring surgery within one hour) while 63% can get surgery within twelve hours. Even in the orthopedic trauma center researched by Bower and Mould [10], the real trauma patients are only 25% of the total non-elective patients. This mismatch is reflected in results that are measuring if the patients are served within their deadline. For instance, Leppäniemi and Jousela [55] show that the highest priority non-electives are usually (in more than 80% of the cases) served within their target time, while this number decreases for the medium and lower priority cases.

Nevertheless, Samudra [77] reports opposite trends, where respectively 74%, 78%, 83% of the three highest urgency categories are served within their deadline.

A second categorization method uses priority scores. To clarify this, it is important to note the distinction between urgency and priority. Urgency is primarily based on medical criteria at the time of arrival of the patient to the hospital. Priority, however, refers to the relative position of the patient with respect to other patients on the waiting list and is often used to develop an admission rule. Priority scores try to take both into account. They are based on a wide variety of aspects often including medical urgency, professional priorities, resource use and time on the waiting list. Mullen [66] provides a comprehensive review of the priority scores developed over time including additive and multiplicative forms. The objective of the priority scores ranges from determining whether a patient is delayed or denied to defining the patient's urgency and importance. In addition, MacCormick et al. [61] published a review on prioritization systems of elective patients in which they discuss the different factors and their weighing. They show that only 13 out of the 50 reviewed studies include recommended waiting times together with a prioritization system.

Testi et al. [88] propose the priority score in Formula (1) where c is the urgency status that is calculated as the weighted sum of the numerical values of three clinical criteria. The same authors compare this scoring algorithm, that determines a relative priority for each patient in the waiting list, to a clinical assessment with a recommended maximum waiting time [88]. They use the need-adjusted-waiting days as a performance measure to include both urgency and priority and conclude that both methods should be used simultaneously. Indeed, Sobolev et al. [81] show that less urgent patients sometimes have a higher probability of admission compared to more urgent cases if a classification system based on urgency is used. This is possibly caused by hospital-related or patient-related delays. Depending on the hospital, priority scores can be used either as a pure classification system or for prioritization purposes.

$$Priority\ score = c * Waiting\ Time \tag{1}$$

Sporadically, multiple non-elective patients arrive approximately at the same time so that next to categorization, priorities need to be set to organize the surgical process. Both Guerriero and Guido [30] as well as Min and Yih [65] argue that a surgery schedule should consider patient priority and that inadequate consideration of patient priority may result in a suboptimal and ineffective schedule. In general, non-elective patients get priority over the patients in the elective schedule and often it is the attending staff surgeon who decides on the exact order. Although several authors ([9,34,70,96]) explicitly mention that the non-electives are served according to the applied classification system (i.e., there is a fixed priority between the different categories), most papers do not mention the prioritization system. Table 7 also shows that first-come, first-served (FCFS) is assumed in several models and simulations, even though the survey of Cardoen et al. [14] reports that about 68% of the respondents indicate that the arrival sequence is a less important factor in determining the priorities. Persson and Persson [72] combine FCFS and medical urgency and add a stand-by system for the emergency staff. Finally, a hospital might prioritize based on a rule or algorithm. For instance, Dexter et al. [18] proposes to prioritize urgent surgeries in increasing order of expected surgery durations if the scheduling objective is to minimize the average length of time each patient waits next to the FCFS and medical priority systems.

Table 7: Categorization based on the prioritization rule

Prioritization rule	Reference
FCFS	[1,23,24,72,93]
According to the classification system	[9,34,70,72,96]
According to an algorithm or rule (e.g., best fit)	[10,18,19]

4.6. Patient volume

In order to 'schedule' non-electives and trade-off capacity between non-electives and electives, it is useful to know the volume of the non-electives or the probability that a non-elective patient will arrive. Since non-electives might come from the ED, it is interesting to first look at the admission rates for ED patients to predict the flow of incoming non-elective patients. Peck et al. [71] studied four hospitals and the admission rate varied from 26% to 32% of the approximate monthly value. However, no daily or hourly rates are provided.

Secondly, another informative number is the percentage of patients that receive surgery and entered the OR as a non-elective. Unfortunately, as shown in Table 8, only a few papers discuss the ratio between electives and non-electives, although this ratio is important in deciding whether or not to dedicate capacity. In addition, the ratio varies widely from hospital to hospital.

Moreover, add-on cases, although differently defined by different authors, include a significant part of the treated patients. For instance, in the case hospital of Dexter et al. [20], where an add-on case is defined as a patient that is scheduled after 7 PM for the next day, 24% of the slots (i.e., a combination of an OR and a date) contain add-on cases. In addition, at least half of the ORs have the last case scheduled or changed within two days of surgery.

Table 8: Number of non-elective patients as a percentage of the total admitted patients

Category	**%**	**Reference**
Non-elective	10-15%	[75]
	14%	[24,93]
	17%	[87]
	20%	[1]
	22%	[77]
	25%	[60]
	33%	[78]
Semi-urgent	40%	[96]
Add-on	6%	[17]
	20%	[34]
	24%[a]	[20]

Note. Papers using the same case hospital are only mentioned once.
[a] 24% of the OR date combinations had at least one add-on case.

5. Flexible policy

In a flexible policy there is one pool of ORs in which all patients, both non-elective as well as elective, are operated.

As explained earlier in Figure 1, a flexible policy contains two options. A first option, used by many hospitals, consists of filling the OR for a given fraction of the full capacity (e.g., 85%) to leave some safety margin or slack for unexpected events (e.g., arrival of non-electives). Clearly, the safety margin decreases the OR capacity which is assigned to electives.

The slack is mostly planned 'virtually' at the end of the day. It can be either a stochastic or a deterministic amount. In the deterministic case, the slack is for instance based on the expected non-elective arrivals. In the papers advocating the stochastic approach, the non-electives are usually handled as a separate, aggregated category, without specifying the different urgency levels. This approach is mostly preferred in papers using a mathematical optimization model covering the master surgery scheduling problem.

The other option is to schedule specific moments throughout the day at which non-electives can enter the elective schedule. These moments can be either break-in-moments (BIMs) [23] or time intervals or breaks. The idea of BIMs is to minimize the time that an arriving non-elective patient has to wait before receiving surgery. Note that no capacity is left open in this case, only a possibility for entering the schedule is created. Alternatively, inserting breaks does leave capacity free at predefined spots in the schedule.

According to a survey of Cardoen et al [14], most hospitals (85%) in Flanders adopt the flexible strategy and plan non-electives in the first OR that becomes available. With regards to urgencies, the majority of the respondents incorporated these patients in the regular program of the appropriate discipline during the day (54%). Another 30% uses the practice of operating the urgent surgeries at the end of the day program and the remaining 16% combines both practices.

The literature on both the required slack as well as on inserting breaks is scarce and these topics form an area for future research.

5.1. Results

The advocates of the flexible policy mainly report on five performance measures as reflected in Table 9 and most papers consider more than one. The results on staff overtime are contradictory. Some authors report a possible increase due to the increased variability [24] in the schedule while others note a reduction in the overtime [93]. In addition, Wullink et al. [93] show that the average overtime per day decreases, but the average number of ORs with overtime per day is (slightly) higher in the flexible strategy. On the contrary, Ferrand et al. [24] indicate that although the average overtime increases, the number of patients served in overtime remains the same. Interestingly, they suggest that in a highly variable system, going towards a more flexible system might benefit both patient categories.

The elective patient waiting time increases according to most authors, because the elective schedule is disturbed by the incoming non-electives. According to the papers advocating a flexible approach, pursuing a more flexible strategy results in a lower waiting time for non-electives. The reasoning behind this is that the non-elective patients have more possibilities to enter the ORs since they can access all ORs. However, proponents of the dedicated strategy argue that the opposite is true, as discussed in section 6. Note that waiting time measures the time between the request for surgery and the moment of surgery and thus indicates the access time.

Furthermore, the flexible policy results in an improved overall utilization and in an increased number of cancellations of the scheduled cases. Surprisingly, the cost of cancellation is overlooked in many papers which might give a biased view on the results of the flexible policy. After all, cancellations are the other side of the trade-off between utilization and overtime. Moreover, since electives are usually admitted to a ward before having surgery, the non-elective arrivals mainly cause inconvenience for the electives, rather than disturbing the processes in the OR [93].

Only a few papers make an enhanced comparison between a flexible and a dedicated policy (e.g., [24,52,87,93]). In section 3, they are classified under both policies, while in sections 4 and 5 they are classified only in the section where they are assessed to be most relevant. The direction of the performance measures when moving towards a more flexible policy, shown in Table 9, is based on those studies or on explicit notes on the direction in other papers. The same reasoning holds for the dedicated policy in section 5.

Table 9: Performance measures in a flexible policy

Performance measure	**Increase/decrease**	**Reference**
Waiting time electives	↑	[1,24] [a]
Waiting time non-electives	↓	[23,24,93]
Staff overtime	No consensus	[1,22,24,32,36,47,48,65,68,73,76,93,96]
Utilization of OR (overall)	↑	[1,19,24,32,36,46,50,93,96]
Cancellations/rescheduling	↑	[1,22,65,84,96]

[a] Adan et al. [1] show that more flexibility decreases the overall waiting time, but their setting focuses on flexibility with regards to scheduling electives.

In addition, other performance measures show up in the reviewed papers. As such, Dhupar et al. [21] show that delayed OR availability for urgent surgeries, which is the highest during regular operating hours, significantly increases the total hospital costs. They provide a retrospective study of five years of data for the appendectomy procedure. Some authors consider the waiting time for both patient groups as one performance measure (e.g., [1]). Also the leveling of resources (e.g., [1,23]), deferrals (e.g., [1]), the financial contributions (e.g.,[46]) and OR undertime (e.g., [96]) are considered.

In the flexible approach of reserving slack, the main goal seems the trade-off between utilization and overtime, while for the break-in-moments non-elective waiting time is considered as main performance measure. Interestingly, van der Lans [52] argue that a higher process variability leads to lower non-elective waiting time in BIM optimization.

A practical drawback of the flexible policy is that a successful implementation requires everyone's collaboration since everyone has to reserve free capacity.

5.2. How are non-electives handled

In general, non-electives are operated on as close as possible to their arrival time. Non-electives arriving during the day who could not be fitted into the schedule are operated on in overtime or during the night and evening shifts. Lovett et al. [60] report that 53% of the non-electives are served outside regular hours (i.e., after 5 PM in the case hospital). Besides, non-electives are often served at the end of the day, after the elective schedule is finished [79,90]. The next subsections discuss the literature on the two options of the flexible policy.

Operational scheduling (option 1)

Building on the results of Wullink et al. [93], van Essen et al. [23] explore the option of break-in-moments for non-electives. They want to minimize the time between the BIMs. This leads to sequencing the surgeries in their assigned OR such that the maximum interval between two consecutive BIMs is minimized. The non-elective waiting time can be reduced by spreading the BIMs as evenly as possible over the day. For an instance with sixteen ORs, the percentage of non-electives that have to wait longer than 30 minutes is smaller than 0.5%. Several constructive and improvement heuristics are applied and tested.

Other authors apply standard scheduling techniques and adapt them to the healthcare context such as Pham and Klinkert [73], who model the problem as a multi-mode blocking job shop problem and develop a mixed integer programming (MIP) formulation. They model the problem of inserting non-electives as the job insertion problem. They insert the non-elective patients such that the resource assignments and the sequence remain the same.

Several hospitals also use heuristics to 'plan' or insert the non-electives. For instance, Azari-Rad et al. [5] report that non-electives who need surgery within two hours of their arrival are assigned to the first available OR. If the arriving non-elective needs to be served between two and eight hours from the arrival time, the patient is assigned to the end of the day (until 11 PM) or is served as the first case of the next morning.

Instead of inserting the non-electives directly into the schedule, another approach is to assign 'buffers' in which the non-electives can be served if needed. This is different from reserving a certain amount of slack (e.g., 10% of capacity) since the buffers are spread out over the ORs and over time (see Figure 1, option 1) and can be variable in size. These buffers protect against unforeseen non-electives, but can also protect against duration variation. In the search for the best spots in the schedule to insert breaks, the research of Klassen and Rohleder [43,44] can provide insights. They study outpatient appointment scheduling and conclude that leaving open slots for urgent patients at the end of the day improves both the percentage of the urgent customers served and the server idle time while open slots at the beginning of the day decrease customer waiting time, but also decrease the percentage of urgent customers served. Moreover, they argue that it is best to evenly spread the open slots over the day, which is similar to the results on BIM. These findings could be tested in an inpatient surgery scheduling environment in future research.

Hans et al. [32] develop a robust surgery schedule in order to ensure maximum OR utilization while minimizing the risk of overtime and cancelled patients. They use a combination of constructive heuristics and local search techniques. This paper can be considered as a mix of option 1 and option 2 of the flexible strategy since the authors use slack without scheduling the slack, but the slack is dependent on the characteristics of the individual surgeries and on the amount of surgeries that is already planned in the OR. None of the previous authors discuss how exactly to include the non-electives into the schedule or into the buffers. This topic and the topic on the impact of inserting breaks into the schedule, form a possible area for future research.

Furthermore, the scheduling of add-on cases has been researched by some authors (e.g., [19,95]). Dexter et al. [19] evaluate ten scheduling algorithms, which are online and offline variants of the best fit and worst fit algorithms, to schedule add-on cases in the 'remaining' OR time. The algorithms are suited for solving the variable-sized bin packing problem with bounded space. Their result is applicable to ORs that schedule one or zero add-on cases per OR. In the offline algorithms the add-on elective cases are batched at a specific cut-off time (e.g., 5 PM the day before surgery). In the case hospitals (one covering 22 inpatient ORs and the other one having six outpatient ORs) there was on average more than an hour remaining in each OR per day. In about 45% of the ORs there was no remaining time. By scheduling the add-on cases, the utilization in both hospitals increased from around 84% to about 93%.

An example of introducing flexibility into the scheduling process is provided by Adan et al. [1]. They combine slack planning (see infra) with different levels of flexibility to schedule electives in the assigned capacity. An example of a flexibility rule is to plan extra patients from the waiting list in case there are fewer patients scheduled than the number that was planned for. Furthermore, the following daily scheduling algorithm is used for dealing with non-electives. Non-electives that arrive during night time are operated on during the night shift and the beginning of the day shift if necessary. If they arrive during the day, they are operated on during the day (after the elective program) and at the beginning of the night shift (if needed). For each arriving non-elective, the decision rules for cancellation (i.e., deferral) are followed. The cancellation decision is based on the estimated required resources at the beginning of the day. Electives that are 'cancelled' are postponed to the same slot in the next week.

Going one step further at the operational level leads us to online (re)scheduling where scheduling happens upon the event of an arrival of a patient. For instance, Erdem et al. [22] develop a genetic algorithm to minimize the cost incurred due to disruptions of non-elective arrivals. Decisions that need to be made are whether or not to admit or defer the non-elective patient and how to adapt the elective schedule upon the arrival of the non-elective patient. They take the cost of postponing or preponing as well as the cost of turning down a patient into account. Important to note is that they deal with one non-elective arrival per day at the beginning of the day. Another recent example is provided by Stuart and Kozan [84], who re-optimize the sequence of surgeries to incorporate the incoming non-electives by modeling a single-machine scheduling problem with due dates. Other approaches that are common for online rescheduling are right shift scheduling, partial rescheduling and re-optimization [92].

In general, however, the literature on how to adapt the elective schedule upon the arrival of non-electives is scarce and constitutes a great area for future research. Similar concepts can be found in other domains. For instance, in proactive-reactive project scheduling, a baseline schedule is created and a reactive policy is applied when during project execution disruptions such as longer activity durations or the unavailability of resources occur (e.g., [15,35]). Moreover, Artigues et al. [2] developed insertion techniques for the static and dynamic resource-constrained project scheduling problem. In the ED context, Lee [54] provides an OR-predictive analytic decisions framework that combines simulation optimization, machine learning and predictive analytics

Capacity (option 2)

The second option is to reserve a specific amount of capacity without scheduling this capacity into the schedule. Van Houdenhoven et al. [37] research a university hospital applying block scheduling, where the schedule is made two weeks ahead. Each surgical department has a different target utilization to account for non-electives and duration variability and so each has 'planned slack'. The authors use the portfolio effect, which states that portfolio risk decreases with increasing diversity (i.e., no correlation between components exists), and cluster the surgical cases with similar variances. This results in less total planned slack. The slack equals the standard deviation of the total duration per block times the risk factor for overtime (e.g., 30%). The optimal slack time depends on the case mix and the block length. For instance, a slack time of 40 min would be optimal for ophthalmology while for ear, nose and throat this would be 110 min. Wullink et al. [93] also reserve slack in each OR and distribute the total amount of slack evenly over the ORs. Each specialty reserves one emergency surgeon per day (e.g., on a day research or administrative tasks are planned for this surgeon). Unfortunately, they do not consider the waiting time of elective patients. Van Houdenhoven et al. [36] calculate the reserved capacity per OR as the average non-elective duration plus the variability in the non-elective durations plus the variability in the elective durations. A norm utilization is calculated per surgical department to address the association between OR utilization, case mix and accepted risk of overtime. Alternatively, Kuo et al. [46] develop a linear programming model where the capacity for non-electives is assumed to be constant among all surgeons.

Adan et al. [1] also introduce slack planning (and flexibility rules) to cope with the deviations between the actual and average flow of patients. They provide a two-stage approach for the master surgery planning, including a MIP and a scheduling algorithm. They calculate the reserved slack based on the arrival rate of non-electives on a specific day and the probability that the non-elective arrives at daytime. Moreover, since the actual number of arriving patients is not equal to the average number of patients in the tactical plan, a number of patients that is higher than the average is scheduled in the tactical plan to create slack in the operational plan. The planned slots are then filled up in the operational phase with or without following several flexibility rules. Afterwards, the execution of non-elective and elective surgeries happens according to the daily scheduling algorithm described earlier in the operational section. They clearly show the trade-off between hospital efficiency and patient satisfaction (e.g., waiting time, cancellations) along the efficient frontier.

Zonderland et al. [96] examine the trade-off between societal cost (due to patient cancellation and waiting time) and the required capacity. They assume that urgent patients are served in a separate OR and focus on fitting the semi-urgent patients into the elective schedule. The authors investigate reservation schemes for semi-urgent patients. A queueing model is used to calculate the fixed OR time that is reserved and a Markov model is used to determine the number of slots for the two-week semi-urgencies to plan in week one. If a one-week semi-urgent patient arrives, first the reserved one-week semi-urgent OR time is used. Then, if this time is not sufficient, electives are cancelled and only then overtime is used. Cancelled electives become semi-urgent patients, which need to be served within the following two weeks. The authors argue that the number of cancellations might increase if only the average behavior of the system is taken into account.

The amount of slack can also be represented by a stochastic variable. Lamiri and Xie [47–50] provide an elective surgery planning including uncertainties in the form of random variables for surgery durations and for the required capacity used by non-elective arrivals. They say that tackling duration variability by introducing slack assumes that the sum of the durations is normally distributed. They formulate the planning problem as a stochastic mathematical program and solve it by first approximating the MIP formulation by the Monte Carlo sampling technique. Later, they extend the model by applying column generation, where each column represents a possible assignment of elective patients to an OR. Heuristics are used to derive a feasible solution to the integer problem, which is then improved by local search. Note that they not include waiting time explicitly, but the models penalize for delaying the scheduling of electives. Also Min and Yih [65] incorporate the capacity used by non-electives as a random variable in a stochastic MIP model. Furthermore, Gerchak et al. [28] looks for the optimal amount of capacity to reserve for non-elective patients each day and shows that this amount is a decreasing function of the number of elective patients waiting.

6. Dedicated policy

In the dedicated policy, a subset of ORs is dedicated to serve a specific group of patients. Most authors try to keep the same total amount of capacity as used or as would be used in the flexible policy. A DOR fits into the idea of reducing flow variability. Often long waiting times are due to a flow problem and not to a resource problem [33] and separating electives and non-electives is one way to reduce the flow variability. By separating the inherent variability from unscheduled emergency cases, the use of elective ORs can be maximized. The purpose of a DOR is often to improve access to care for both elective and non-elective patients and to reduce rescheduling and cancellation actions.

From the perspective of the elective patients, capacity is reduced in both the flexible and the dedicated scenario. In the former, emergency disruptions indirectly reduce capacity [24] while in the latter the capacity is directly reduced. The main question is which reduction is the best option.

When dedicating capacity, it is important to decide for which patient categories the capacity will be reserved. Although the goal is mostly to reserve capacity for non-electives (e.g.,[24,34,55,87]), the access to the DOR might be limited to trauma orthopedic patients (e.g., [10]), to urgent and semi-urgent patients (e.g., [9]), to add-on cases or to a combination of patient types (e.g., [75,80]).

A special case of DORs is when almost all ORs are dedicated to non-elective arrivals. This happens for instance in case of a disaster, where many rooms need to be cleared to provide timely access for incoming life-threatening cases.

6.1. Results

When using DORs, a clear trade-off between the reduction in flexibility and the increase in access time appears. In other words, it is the trade-off between a decrease in the waiting lists and the reduction of cancellations and overtime. This policy is in line with the recent guidelines published by the Royal College of Surgeons to separate both patients flows [85] and the idea of reducing system-wide variability by separating the flows [33,58].

In general, papers discussing the impact of DORs only report a very limited set of performance measures, which makes it difficult to properly assess the full range of the impact. The performance measures that are most used are the non-elective waiting time, the staff overtime, the OR utilization and the disruptions, as shown in Table 10. Note that comparing the results of the patient waiting time (for both categories) and the staff overtime for the dedicated policy with the results of the flexible policy (Table 9) are contradictory.

Firstly, the impact on the staff overtime is positive. However, many of the papers report mainly on a decrease in the number of non-electives served in overtime, but fail to report on the overall impact on overtime, which includes both the overtime of the electives and non-electives as well as the range of regular hours. This lack of information skews the interpretation of the results significantly.

Table 10: Performance measures in a dedicated policy

Performance measure	Increase/decrease	Reference
Non-elective waiting time	↓	[9,34,70,75,78,80][b]
Staff overtime	↓	[7,9,34,55,60,72,74,75,78,80]
Utilization of regular OR	↑	[24,78,80]
Utilization of DOR	[37%, 42%, 53%, 60%, 62%, 85%]	[7][75][34][80][78][55][c]
Disruptions[a] (electives)	↓	[9,55,72,80]
Elective patient waiting time	↑	[72]
Throughput	↑	[72,75,78,80]

[a] Disruption includes the number of disruptions, the number of rescheduled patients and the number of cancellations.
[b] [34] report only the waiting time for priority three patients.
[c] References are in the order of the (increasing) utilization rates.

Secondly, the utilization of the regular (elective) ORs is higher thanks to the reduction in disruptions. However, the utilization of the DORs ranges from 24% to 85% and depends on several factors like case mix, patient volume and scheduling policies. In contrast to the other authors, Sandbaek et al. [78] show that the OR utilization for both the dedicated ORs and the elective ORs is higher than in the flexible policy. However, since they implement also a new classification and booking system, this effect might be caused by other factors than the dedicated policy.

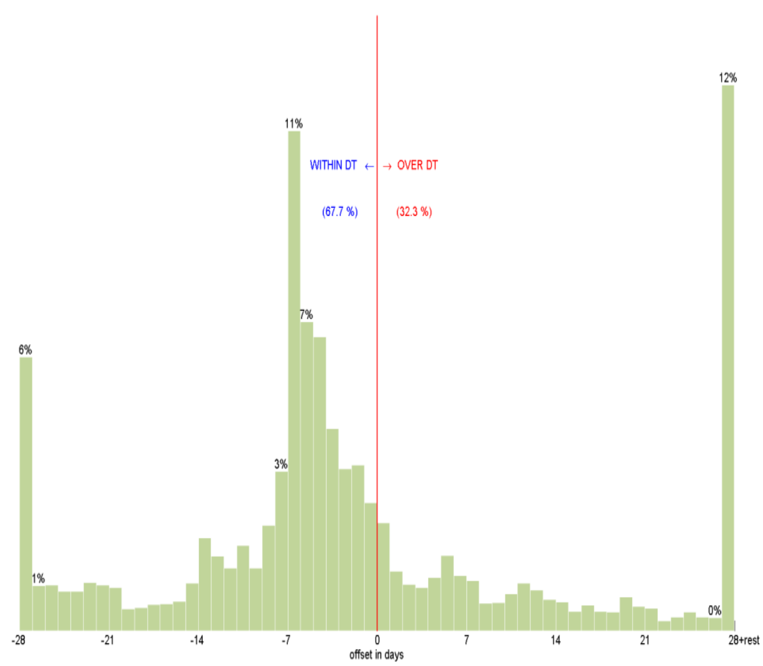
Finally, splitting the flow of non-electives from the one of electives causes fewer unscheduled events in the elective schedule, which is consistent with the results from the flexible policy. This means that fewer electives must be cancelled due to the arrival of non-electives and fewer rescheduling actions (to another OR) must take place.

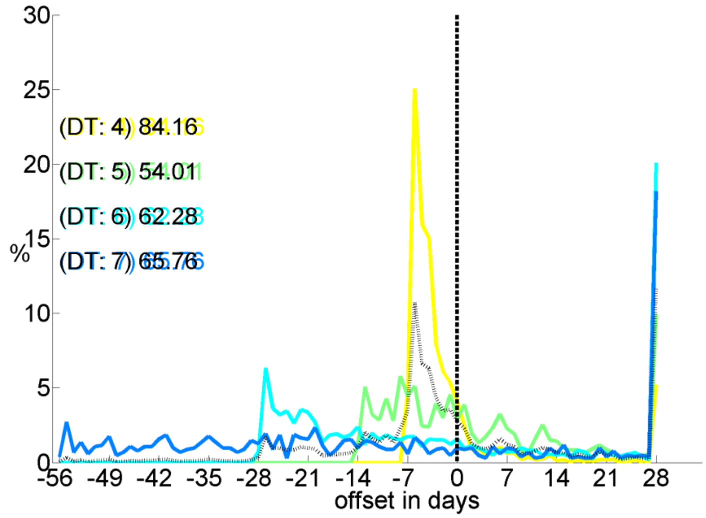
Concerning the waiting time results several remarks need to be addressed, that contribute to the difficulties in comparing or assessing the results of the papers on dedicated capacity. First, the waiting time for both elective as well as non-elective patients is discussed in a surprisingly low number of papers. To assess the impact of introducing dedicated capacity, both performance measures need to be quantified. Similarly, the number of papers that clearly mention all elements of the trade-off between utilization, the number of DORs and the waiting time for all patient categories is limited.

Secondly, a commonly used measure is the average waiting time over all patients. This measure is sensitive to outliers and to differences between the patient categories. An additional element that should accompany this measure is the distribution of the waiting time. Note that for the scheduled patients the waiting time can be split into direct and indirect waiting time [16].

A third remark relates to the hospitals working with a classification system that prescribes a waiting time target for each category. In this case, the average waiting time does not provide the right information. After all, waiting for an average of twelve hours might be acceptable and even desired for some patients. Therefore, it might be more insightful to know whether or not a patient is served within his/her waiting time target, regardless of the total waiting time. Analogous to the second remark, unfortunately only a few of the papers using a target time window mention the distribution of the deviations from the target. This information would give a deeper understanding of the performance of the system with regards to waiting time and how well the hospital serves each patient group. This might be relevant when equitable access is pursued by the hospital. As an example, Figure 4 shows the deviation in days from the due time target and provides a clear view on the percentage of patients that are served outside the interval and on the distribution of the offset. Also Sandbaek et al. [78] provide information on both the median waiting time and the proportion of patients within their target time. Interestingly, the results differ for the three non-elective categories.

Figure 4: Deviation from the due time target in days for all patients (top) and for the elective categories (bottom)[77]





Finally, the waiting time for non-elective patients greatly depends on the exact implementation of the policy. For instance, if only one room is dedicated out of twenty ORs, the waiting time for non-electives will probably increase drastically (considering the patient volumes in section 3.6). After all, they now only have a small part of the capacity, which the patients access in a highly variable way, resulting in a queue for the non-electives. However, dedicating more rooms in the same setting might have a positive effect on the waiting time of the non-electives.

Looking for new performance measures that take the differences between the patient categories into account is an area for future research.

A practical drawback of this policy is that if the emergency staff is sent away to deal with staff shortages in the elective ORs, the 'emergency' team might be incomplete upon arrival of an emergency [93].

6.2. How are non-electives handled

In the papers that focus on only a few departments (see Table 2), several authors discuss the option of dedicating an OR to orthopedic (trauma) patients. As such, Bhattacharyya et al. [9] and Bower and Mould [10] investigate the effect of a dedicated orthopedic trauma OR since the orthopedic trauma cases are often waitlisted and served in overtime. Bhattacharyya et al. [9] devote one OR to less urgent orthopedic cases. The room is controlled by orthopedics to schedule cases in a priority order determined by the day's attending staff surgeon. Results show a significant shift to performing non-electives during daytime. This reduces complications since night time cases last longer due to the sometimes difficult surgeries with inexperienced staff on odd hours. The analysis only focusses on two common surgical cases. Surprisingly, they report a utilization of 88% for the

DOR, while the average for all services is 80%. However, in their case study still 73% of the left-over add-on hours (i.e., the ones that are not performed in the trauma room) are scheduled outside the blocks and performed in overtime; the others are placed in scheduled blocks that would have gone underutilized.

Persson and Persson [72] also focus on the problem of how to schedule elective and non-elective cases at an orthopedic department. In the orthopedic department, the non-elective waiting time is more flexible since most patients can wait 24 hours before getting surgery. They model an optimization model and incorporate it in a simulation model. They introduce stand-by patients who can be called upon (on pre-defined dates that fit for the patient) when there is a free spot. The authors examine a setting with only two ORs. They propose a new setting in which the second OR is fully assigned to non-elective surgeries and stand-by patients. The optimization model consists of a bin packing model that minimizes the cost of surgery, including a non-linear, non-decreasing cost related to the waiting time and costs with regards to cancellation and overtime. The results show a significant decrease in the number of surgery cancellations and overtime at the expense of an increase in waiting time for the electives. The throughput remains similar.

Heng and Wright [34] examine the installation of a dedicated room for add-on cases. They look at the delayed electives, which are the electives who experience a delay of more than 30 minutes if a non-elective was added to the OR, and at the overtime when a non-elective was added. They show that the lowest priority non-electives, priority 3 patients who need surgery within twelve hours, experience less waiting time, are served more within their target waiting levels and are operated on less in evening and night shifts. The DOR has little effect on access for priority 1 and 2 patients. General Surgery, Orthopedics and Neurosurgery make the most use of the room and there would be insufficient volume for a service-specific DOR. The DOR leads to fewer elective cancellations and less overrun minutes in the elective rooms. Surprisingly, the number of delayed electives due to add-on cases remained similar after the introduction of an add-on room.

In Paul and MacDonald [70] all the patients (from all departments) receive a category-specific waiting time target and can transition to a higher priority class if they have to wait extensively. The authors develop a non-preemptive, multi-priority queuing model to manage the non-elective surgeries. They use both queuing theory and DES to determine the required number of DORs. They also present algorithms that estimate the appropriate pricing for the surgeries, differentiated by priority level and given the patient demand and the resources reserved to meet this demand. The price is set such that a patient who waits longer due to a higher priority patient is charged less. They show that the survivability rate (i.e., the percentage of patients within their waiting limit) does not depend on the patient mix between non-electives or on the surgery duration variations within the total volume. However, they assume that all surgeries follow the same duration distribution. They use a probit regression to determine the relation between the required number of ORs, and the total patient volume and the average surgery time. Next, they determine the relation between the proportion of patients of each priority exceeding their waiting time target and the optimal number of ORs, the average surgery time and the total patient volume. They also use transforms and the resulting mathematical moments for the case of one OR.

On a smaller scale, Lovet and Katchburian [60] show that the introduction of an afternoon emergency theater list, which is coordinated by a consultant anesthetist, drastically improved the number of non-electives performed during daytime. Although the number of patients served between 5 PM and 10 PM remained quite high, the reduction in the number of patients served after 10 PM is significant. Similarly, Barlow et al. [7] investigate the effect of an afternoon OR for emergency surgeries of a mix of specialties on the number of operations performed during night-time. For general surgery there was a significant reduction. However, utilization of the DOR was only 37% and only 36% of the non-elective operations were performed in the afternoon session while the majority was still performed in overtime. Furthermore, Leppäniemi and Jousela [55] show that a classification system with three categories for non-electives, combined with a DOR decreases the number of non-electives performed at night and the waiting time for non-electives.

Recently, Smith et al. [80] focused on the identification, quantification and the elimination of artificial variability in order optimally manage the flow of elective patients. They separate the patient flows that cause the natural and artificial variation and test different room allocations for elective and non-elective patients. As such, they separate emergent and urgent cases, work-in cases, (who need to be served within one week) and elective cases. Making a different choice (e.g., dedicating three rooms to all urgent and emergent surgeries), results in significantly different performance (e.g., lower average OR utilization). Subsequently, they allocate sufficient block time for 125% of each service's current demand in the elective rooms in order that, ceteris paribus, 80% of the allocated (elective) block time would be used. Next, they assign block time in order to smoothen the volumes throughout the week. In addition to the results reported in Table 10, they report a significant decrease in the day-to-day variability and an increase in net operating income.

Finally, Ferrand et al. [24] developed a simulation to compare the flexible and focused resource allocation policies. They assume that the elective schedule is fixed and patients arrive in batches according to fixed inter-arrival times. In the scenario with dedicated rooms, the arrival pattern and the elective schedule is assumed to be the same as for the flexible policy, resulting in assigning more elective patients per elective room. However, the utilization in the DORs covers a wide range from 24% to 75%. They collect the average and the maximum patient waiting times as well as the proportion of non-elective patients waiting longer than 30 minutes. Even if the volume of non-electives is increased by three standard deviations, the waiting time of both patient categories is still balanced by the policy of dedicating five out of twenty ORs to non-electives. In the flexible policy, the maximum waiting time for emergencies is 22 minutes, while in the focused policy this number rises to 191 minutes. In addition, 30% of the non-electives waited for more than 30 minutes (target). The authors show that the reduced OR hours allocated to electives are more than compensated by the elimination of emergency disruptions in the flexible policy. However, the dedicated policy increases the waiting time for non-electives, but the size of the effect is dependent on the variability in the processing times where higher variability results in lower waiting times. Interestingly, in the two scenarios with varying coefficients of variation, the average waiting time of emergencies goes down in both policies when processing time variability increases. However, the maximum waiting time for emergencies increases for the dedicated policy and decreases for the flexible policy.

6.3. Dedicated team

Not just physical resources must be dedicated to a specific patient group, but also human resources. This aspect is often overlooked in the literature dealing with operational OR planning. In other words, staff is often assumed to be available when constructing the schedule.

Generally, an on-call team including surgeons, anesthesiologists and nurses is available to assist in case of emergency needs. However, the reviewed papers include a few specific suggestions regarding a dedicated surgical staff. Bhattacharyya et al. [9] describe a system where one staff orthopedic surgeon is designated to cover each day of the week in the dedicated orthopedic trauma OR. Still, flexibility in switching surgeons for a case is desirable. Surgeons usually are on call the night before their designated trauma room days. Heng and Wright [34] propose a similar system. Another possibility is to have a special team that is experienced in the most common emergency procedures over all disciplines [45]. If no non-electives arrive, this team is used for handling semi-elective procedures or replacing sick colleagues. Furthermore, Ryckman et al. [75] make a distinction between divisions that experience frequent urgent needs where a surgeon is assigned to the non-elective cases and divisions with fewer urgent patients where an on-call system is used. This on-call system is also proposed by Persson and Persson [72] for weekend shifts.

Finally, van Oostrum et al. [68] determine an optimal OR team composition for operating during night-time with the two-fold goal of minimizing staffing costs and providing care within the safety interval. They show that modeling safety intervals can reduce the required staff while still ensuring good quality of care.

6.4. How many ORs to dedicate

The answer to this question does not seem to be straightforward. The proportion of ORs that are dedicated (in most cases to non-electives) varies widely and the same holds for the applied methods. Additionally, the patient categories to which the capacity is dedicated also differ.

Wullink et al. [93] consider one DOR out of a total of twelve staffed ORs for an average of five non-elective patients per day (mean case duration of 126 minutes). The other eleven ORs serve on average 32 electives (mean case duration of 142 minutes). On the contrary, Leppäniemi and Jouseia reserve two extra DORs for non-electives for the same total amount of ORs. Another example of diverse allocation mechanisms is provided by Tancrez et al. [87] and Zhang et al. [94] who reserve both only one room out of respectively six and twenty ORs.

In order to get a clearer view on the right amount of dedicated resources, Ferrand et al. [24] test various divisions of the twenty ORs between the two patient groups and trade off elective waiting time with non-elective waiting time. They opt for five DORs. An analysis with the average number of overtime patients and the average and maximum overtime gives the same results. They also show that with an increasing number of DORs (ranging from three to seven), the average elective waiting time and the number of patients served in overtime increases and the average elective waiting time decreases. Interestingly, the overtime first decreases and then increases from five DORs onwards. Besides, when dedicating three ORs, the percentage of patients outside the target waiting time is higher for non-electives than for electives. This switches from four ORs onwards.

Most authors use DES and queueing theory to find the amount of DORs. For instance, Ferrand et al. [24] argue that a steady state queuing analysis depicts at least four DORs, but this analysis does not take overtime into account. Other authors [52,78] calculate the number of DORs by dividing the total expected capacity required for urgent surgery by the capacity per OR and take the smallest following integer. The 'last' OR is then filled up with electives for the remainder of its capacity. Paul and MacDonald [70] calculate the required number of ORs based on the total patient volume and the average surgery time in order to meet the 5% threshold for the survivability for all patients.

Smith et al. [80] not only dedicate rooms (2) to urgent and emergent cases (including heart and lung transplants) and prior night emergencies, but also to work-in cases of predefined specialties and abdominal transplant (2 rooms) and to electives (15 rooms). Ryckman et al. [75] test different settings and propose a similar division for twenty rooms. Steins et al. [82] dedicate two out of eighteen ORs to non-electives and dedicate some other ORs to a specific type of surgery like eye and oral surgery. Likewise, the outpatients and inpatients ORs are separated. Additionally, they separate the streams for weekdays, weekends and nights.

Interestingly, some authors show that adapting the opening hours results in better system performance (e.g., [72]). As such, Smith et al. [80] propose to extend the opening hours of the OR for work-in cases by two hours (compared to the regular working hours) and the emergency ORs in [82] are available 24 hours a day, seven days a week. Additionally, Steins et al. [82] show that the opening hours played a key role in the reduced cancellations and the increased utilization.

7. Hybrid policy

The hybrid policy consists of a mix of dedicated and flexible resources. As such, some rooms can be mainly allocated to one patient group, but still be used by other categories under specific conditions. This policy tries to obtain a better trade-off between flexibility and access time than the previous two policies.

Interestingly, Gupta et al. [31] confirm this use of a mix of both flexible and dedicated resources in the manufacturing industry. Also Ata and Van Mieghem [3] investigate whether to use dedicated resources or an integrated network for serving two customer classes. They also show that the result depends on the service quality and the demand characteristics. Moreover, they provide the conditions for using each strategy.

In the hybrid policy different topics need to be addressed, which are all researched to a limited extent in OR planning and thus form an opportunity for future research. The number of dedicated and flexible resources must be decided. Next, the rules for accessing the flexible capacity need to be outlined, just like in the regular flexible policy. The difference is, however, that in the hybrid policy a part of the patients (a predefined group) is already served in the dedicated capacity, which might influence the operational rules for the flexible capacity. Another option is to originally dedicate capacity, but allow under specific conditions patients to overflow. Decisions on when and how to overflow are then necessary. The next paragraphs briefly discuss two options.

First, the DOR for non-electives might be used by electives, but clear guidelines on when and how to arrange this are lacking in the literature. One of the limited examples is provided by Bower and Mould [10], who install an orthopedic trauma session each day covering five hours out of the available seven hours. They suggest to schedule the remaining two hours for electives. The elective patients then accept a probability (15%) of being rescheduled in return for an earlier treatment. This results in a mix of dedicated capacity (five hours of trauma sessions) and flexible capacity (planned for electives, but still available for trauma cases if required). The authors show that a significant increase in elective throughput and utilization of the DOR can be achieved by this allocation.

Second, no uniform rules exist for the overflow of non-electives to elective capacity either. The highest priority category might still be served in the OR which is assigned to the non-electives' specialty in the MSS, especially when no DOR is available (e.g., [94]), while the other urgency categories are always served in the DOR (e.g., [82]). Exceptionally, Zhang et al. [94] look at all relevant performance measures and build a MIP to allocate OR capacity to different types of surgeries. Unfortunately, they do not provide a comparison with the scenario without DOR. In addition, long, complicated emergency cases, such as transplant or cardiovascular surgery, might not be suited for a DOR since this will greatly disturb the flow for the non-electives, especially when only one DOR is available. Often a separate room is reserved for these surgeries or the surgeries are done in the regular time of the discipline.

Tancrez et al. [86,87] also research a setting with one DOR and five or six flexible rooms in which non-electives can still enter with priority if the DOR is occupied. They examine the impact of stochasticity in surgery durations, unexpected non-elective arrivals and blocking due to a full recovery unit. Interestingly, most papers on the dedicated policy report a decrease in the overtime while Tancrez et al. [87] report an increase. They argue that a decrease in capacity for the electives comes at the cost of either an increase of overtime or fewer planned operations within an acceptable overtime. The authors extend their first model [86] and include blocking in the OR because of a full recovery unit, together with an illustration on a real hospital case. They use a continuous Markov model. They quantify the disruptions of the schedule by the non-electives and show that the disruptions increase more than proportionally with the number of arriving non-electives. Dedicating an OR decreases the disruption rate, next to the average non-elective waiting time and the probability of being served within 30 minutes.

A recent example of a hybrid policy for a large number of ORs (20) is provided by Ferrand et al. [25]. They examine different configurations of flexible and dedicated rooms and show that this policy outperforms the other two policies in terms of improved waiting time for both patient categories and lower overtime. They provide guidelines for the case where a limited amount of ORs is dedicated and for the case where a limited amount of ORs is made flexible. They also emphasize the importance of incorporating the prioritization among patients for obtaining the results.

Finally, note that releasing OR block time is a special case of (hybrid) OR allocations. The capacity is released under predefined circumstances and the released capacity can be dedicated to a specific group of patients or can be released as flexible capacity. For instance, Bhattacharyya et al. [9] report that the DOR is freed up for other services if nothing is scheduled by 7 AM or if the cases end before 5 PM. Similarly, Persson and Persson [72] introduce stand-by patients, who are ready for having surgery in case of leftover capacity.

8. Conclusions and future research

This paper reviews the trade-offs in the OR planning for elective and non-elective patients at both the tactical and the operational level. These trade-offs are caused due different sources of variability. In general, three policies are pursued to handle the trade-offs: the flexible, dedicated and hybrid policy. They differ in the amount of capacity that is dedicated to a certain patient category. The majority of the research focuses on the flexible and the dedicated policy. The hybrid policy only received more attention recently and is a possible area for future research.

Most hospitals want to ensure acceptable access time while still keeping the resource utilization high. What acceptable means, depends on the patient classification used in the hospital. For the hospitals including target waiting times a standardized classification is lacking, which reduces the transferability of the results. The impact of the classification system on the results could be addressed in future research.

Unfortunately, the setting of the reviewed papers differs a lot, which renders a comparison difficult. For instance, not all papers look at all departments in the hospital, but focus on one or a few department(s). This raises the question whether different departments might be better suited for different policies.

The main trade-offs include overtime, utilization, schedule disruptions and waiting time for elective and non-elective patient. Pursuing a more flexible policy should lead to fewer schedule disruptions and higher overall utilization. However, the effect on overtime and waiting time for both patient categories is still unclear. Although the performance of all drivers is required to get a complete view on the performance of the policy, often the analysis is limited to only a few of them. In future research, looking for new and more comprehensive performance measures might be useful. Additionally, the relation between the policies and the patient mix, the duration characteristics and the available number of ORs can be further explored.

Going back to the original need of finding policies for handling the trade-offs in the OR, the review shows that this need is not yet fully satisfied. The dedicated and the flexible policy proved to be efficient for certain case hospitals. However, the impact of all the drivers (e.g., waiting time, utilization) on the choice of the policies is researched to a limited extent. In addition, the benefits of hybrid policies are not yet clear. Next to the hybrid policy, the impact of inserting breaks in the schedule forms an area for future research. In general, a clearer view on how to divide capacity and which elements drive such decisions is required. New algorithms or methods could be developed for this purpose in future research.

Acknowledgments

We acknowledge the support given to this research by the Research Foundation – Flanders (FWO-Vlaanderen) as Aspirant (Carla Van Riet).

9. References

[1] Adan I., Bekkers J., Dellaert N., Jeunet J., Vissers J., Improving operational effectiveness of tactical master plans for emergency and elective patients under stochastic demand and capacitated resources. Eur J Oper Res 213(1) (2011) 290–308.

[2] Artigues C., Michelon P., Reusser S., Insertion techniques for static and dynamic resource-constrained project scheduling. Eur J Oper Res 149(2) (2003) 249–67.

[3] Ata B., Van Mieghem J. a., The value of partial resource pooling: Should a service network be integrated or product-focused? Manage Sci 55(1) (2009) 115–31.

[4] Augusto V., Xie X., Grimaud F., A framework for the modeling and simulation of health care systems. In: Proceedings of the 3rd Annual IEEE Conference on Automation Science and Engineering. IEEE, 2007:231–6.

[5] Azari-Rad S., Yontef A.L., Aleman D.M., Urbach D.R., Reducing elective general surgery cancellations at a Canadian hospital. Can J Surg 56(2) (2013) 113–8.

[6] Barkaoui K., Dechambre P., Hachicha R., Verification and optimisation of an operating room workflow. In: Proceedings of the 35th Hawaii International Conference on System Sciences. 2002:1–10.

[7] Barlow A.P., Wilkinson D.A., Wordsworth M., Eyre-Brook I.A., An emergency daytime theatre list: Utilisation and impact on clinical practice. Ann R Coll Surg Engl 75(6) (1993) 441–4.

[8] Beliën J., Demeulemeester E., Building cyclic master surgery schedules with leveled resulting bed occupancy. Eur J Oper Res 176(2) (2007) 1185–204.

[9] Bhattacharyya T., Vrahas M.S., Morrison S.M., Kim E., Wiklund R. a, Smith R.M., Rubash H.E., The value of the dedicated orthopaedic trauma operating room. J Trauma 60(6) (2006) 1336–41.

[10] Bowers J., Mould G., Managing uncertainty in orthopaedic trauma theatres. Eur J Oper Res 154(3) (2004) 599–608.

[11] Brailsford S., Overcoming the barriers to implementation of operations research simulation models in healthcare. Clin Investig Med 28(6) (2005) 312–5.

[12] Brailsford S.C., Harper P.R., Patel B., Pitt M., An analysis of the academic literature on simulation and modelling in health care. J Simul 3(3) (2009) 130–40.

[13] Cardoen B., Demeulemeester E., Beliën J., Operating room planning and scheduling: A literature review. Eur J Oper Res 201(3) (2010) 921–32.

[14] Cardoen B., Demeulemeester E., On the use of planning models in the operating theatre: Results of a survey in Flanders. Int J Health Plann Manage 25(4) (2010) 400–14.

[15] Demeulemeester E., Herroelen W., Leus R., Proactive-reactive project scheduling. In: Artigues C., Demassey S., Nron E., eds. Resource-Constrained Project Scheduling: Models, Algorithms, Extensions and Applications. New Jersey: John Wiley & Sons, 2008:203–11.

[16] Deo S., Iravani S., Jiang T., Smilowitz K., Samuelson S., Improving health outcomes through better capacity allocation in a community-based chronic care model. Oper Res 61(6) (2013) 1277–94.

[17] Dexter F., Macario A., Lubarsky D.A., Burns D.D., Statistical method to evaluate management strategies to decrease variability in operating room utilization. Anesthesiology 91(1) (1999) 262–74.

[18] Dexter F., Macario A., Traub R.D., Optimal sequencing of urgent surgical cases. J Clin Monit Comput 15(3-4) (1999) 153–62.

[19] Dexter F., Macario A., Traub R.D., Which algorithm for scheduling add-on elective cases maximizes operating room utilization? Anesthesiology 91(5) (1999) 1491–500.

[20] Dexter F., Shi P., Epstein R.H., Descriptive study of case scheduling and cancellations within 1 week of the day of surgery. Anesth Analg 115(5) (2012) 1188–95.

[21] Dhupar R., Evankovich J., Klune J.R., Vargas L.G., Hughes S.J., Delayed operating room availability significantly impacts the total hospital costs of an urgent surgical procedure. Surgery 150(2) (2011) 299–305.

[22] Erdem E., Qu X., Shi J., Rescheduling of elective patients upon the arrival of emergency patients. Decis Support Syst 54(1) (2012) 551–63.

[23] Van Essen J.T., Hans E.W., Hurink J.L., Oversberg a., Minimizing the waiting time for emergency surgery. Oper Res Heal Care 1(2-3) (2012) 34–44.

[24] Ferrand Y.B., Magazine M.J., Rao U.S., Comparing two operating-room-allocation policies for elective and emergency surgeries. In: Johansson B., Jain S., Montoya-Torres J., Hugan J., Yüceson E., eds. Proceedings of the 2010 Winter Simulation Conference. IEEE, 2010:2364–74.

[25] Ferrand Y.B., Magazine M.J., Rao U.S., Partially flexible operating rooms for elective and emergency surgeries. Decis Sci 45(5) (2014) 819–47.

[26] Ferrand Y.B., Magazine M.J., Rao U.S., Managing operating room efficiency and responsiveness for emergency and elective surgeries: A literature survey. IIE Trans Healthc Syst Eng 4(1) (2014) 49–64.

[27] Fomundam S., Herrmann J.W., A survey of queuing theory applications in healthcare. 2007. Institute for Systems Research Technical Report No.24 http://drum.lib.umd.edu/handle/1903/7222.

[28] Gerchak Y., Gupta D., Henig M., Reservation planning for elective surgery under uncertain demand for emergency surgery. Manage Sci 42(3) (1996) 321–34.

[29] Green L., Queueing analysis in healthcare. In: Patient Flow: Reducing Delay in Healthcare Delivery. Springer US, 2006:281–307.

[30] Guerriero F., Guido R., Operational research in the management of the operating theatre: A survey. Health Care Manag Sci 14(1) (2011) 89–114.

[31] Gupta D., Gerchak Y., Buzacott J.A., The optimal mix of flexible and dedicated manufacturing capacities: Hedging against demand uncertainty. Int. J. Prod. Econ. 28(3) (1992) 309–19.

[32] Hans E.W., Wullink G., Van Houdenhoven M., Kazemier G., Robust surgery loading. Eur J Oper Res 185(3) (2008) 1038–50.

[33] Haraden C., Resar R., Patient flow in hospitals: Understanding and controlling it better. Front Health Serv Manage 20(4) (2004) 3–15.

[34] Heng M., Wright J.G., Dedicated operating room for emergency surgery improves access and efficiency. Can J Surg 56(3) (2013) 167–74.

[35] Herroelen W., Leus R., Robust and reactive project scheduling: A review and classification of procedures. Int J Prod Res 42(8) (2004) 1599–620.

[36] Van Houdenhoven M., Hans E.W., Klein J., Wullink G., Kazemier G., A norm utilisation for scarce hospital resources: Evidence from operating rooms in a Dutch university hospital. J Med Syst 31(4) (2007) 231–6.

[37] Van Houdenhoven M., van Oostrum J.M., Hans E.W., Wullink G., Kazemier G., Improving operating room efficiency by applying bin-packing and portfolio techniques to surgical case scheduling. Anesth Analg 105(3) (2007) 707–14.

[38] Howard R.A., Comments on the origin and application of Markov decision processes. Oper Res 50(1) (2002) 100–2.

[39] Huschka T.R., Narr B.J., Denton B.T., Thompson A.C., Using simulation in the implementation of an outpatient procedure center. In: Mason S.J., Hill R.R., Mönch L., Rose O., Jefferson T., Fowler J.W., eds. Proceedings of the 2008 Winter Simulation Conference. IEEE, 2008:1547–52.

[40] Jahangirian M., Naseer A., Stergioulas L., Young T., Eldabi T., Brailsford S., Patel B., Harper P., Simulation in health-care: Lessons from other sectors. Oper Res 12(3) (2012) 45–55.

[41] Jun J.B., Jacobson S.H., Swisher J.R., Application of discrete-event simulation in health care clinics: A survey. J Oper Res Soc 50(2) (1999) 109–23.

[42] Katsaliaki K., Mustafee N., Applications of simulation within the healthcare context. J Oper Res Soc 62(8) (2011) 1431–51.

[43] Klassen K.J., Rohleder T.R., Scheduling outpatient appointments in a dynamic environment. J Oper Manag 14(2) (1996) 83–101.

[44] Klassen K.J., Rohleder T.R., Outpatient appointment scheduling with urgent clients in a dynamic, multi-period environment. Int J Serv Ind Manag 15(2) (2004) 167–86.

[45] Klimek M., Van Houdenhoven M., Ottens T., Operating theatres: Organization, costs and audit. In: Gullo A., ed. Anaesthesia, Pain, Intensive Care and Emergency APICE. Springer Milan, 2008:263–8.

[46] Kuo P., Schroeder R., Mahaffey S., Bollinger R., Optimization of operating room allocation using linear programming techniques. J Am Coll Surg 197(6) (2003) 889–95.

[47] Lamiri M., Dreo J., Xie X., Operating room planning with random surgery times. In: Proceedings of the 3rd Annual IEEE Conference on Automation Science and Engineering. IEEE, 2007:521–6.

[48] Lamiri M., Grimaud F., Xie X., Optimization methods for a stochastic surgery planning problem. Int J Prod Econ 120(2) (2009) 400–10.

[49] Lamiri M., Xie X., Dolgui A., Grimaud F., A stochastic model for operating room planning with elective and emergency demand for surgery. Eur J Oper Res 185(3) (2008) 1026–37.

[50] Lamiri M., Xie X., Zhang S., Column generation approach to operating theater planning with elective and emergency patients. IIE Trans 40(9) (2008) 838–52.

[51] Lankester B.J., Paterson M.P., Capon G., Belcher J., Delays in orthopaedic trauma treatment: Setting standards for the time interval between admission and operation. Ann R Coll Surg Engl 82(5) (2000) 322–6.

[52] Van der Lans M., Hans E.W., Hurink J.L., Wullink G., Van Houdenhoven M., Kazemier G., Anticipating urgent surgery in operating room departments. 2005. Working paper No.158.

[53] Lebowitz P., Why can't my procedures start on time? AORN J 77(3) (2003) 594–7.

[54] Lee E.K., Yuan F., Zhou R., Lahlou S., Post E., Wright M.D., Atallah H.Y., System analytics: Modeling and optimizing emergency department workflow. In: Hui Y., Lee E.K., eds. Healthcare Data Analytics. Wiley Series in Operations Research and Management Science. John Wiley & Sons, 2014

[55] Leppäniemi A., Jousela I., A traffic-light coding system to organize emergency surgery across surgical disciplines. Br J Surg 101(1) (2014) 134–40.

[56] Levtzion-Korach O., Murphy K., Madden S., Dempsey C., For urgent and emergent cases, which one goes to the OR first? OR Manager 26(7) (2010) 5–8.

[57] Litvak E., Long M.C., Cooper A., McManus M.L., Emergency department diversion: Causes and solutions. Acad Emerg Med 8(11) (2001) 1107–10.

[58] Litvak E., Long M.C., Cost and quality under managed care: Irreconcilable differences. Am J Manag Care 6(3) (2000) 305–12.

[59] Litvak N., Rijsbergen M., Boucherie R., Van Houdenhoven M., Managing the overflow of intensive care patients. Eur J Oper Res 185(3) (2008) 998–1010.

[60] Lovett B.E., Katchburian M.V., Emergency surgery: Half a day does make a difference. Ann R Coll Surg Engl 81(1) (1999) 62–4.

[61] MacCormick A.D., Collecutt W.G., Parry B.R., Prioritizing patients for elective surgery: A systematic review. ANZ J Surg 73(8) (2003) 633–42.

[62] May J., Spangler W.E., Strum D.P., Vargas L.G., The surgical scheduling problem: Current research and future opportunities. Prod Oper Manag 20(3) (2011) 392–405.

[63] McManus M.L., Long M.C., Cooper A., Litvak E., Queuing theory accurately models the need for critical care resources. Anesthesiology 100(5) (2004) 1271–6.

[64] Mielczarek B., Uzialko-Mydlikowska J., Application of computer simulation modeling in the health care sector: A survey. Simulation 88(2) (2010) 197–216.

[65] Min D., Yih Y., Scheduling elective surgery under uncertainty and downstream capacity constraints. Eur J Oper Res 206(3) (2010) 642–52.

[66] Mullen P.M., Prioritising waiting lists: How and why? Eur J Oper Res 150(1) (2003) 32–45.

[67] Olivares M., Terwiesch C., Cassorla L., Structural Estimation of the Newsvendor Model: An Application to Reserving Operating Room Time. Manage Sci 54(1) (2008) 41–55.

[68] Van Oostrum J.M., Van Houdenhoven M., Vrielink M.M.J., Klein J., Hans E.W., Klimek M., Wullink G., Steyerberg E.W., Kazemier G., A simulation model for determining the optimal size of emergency teams on call in the operating room at night. Anesth Analg 107(5) (2008) 1655–62.

[69] Patrick J., Puterman M., Reducing wait times through operations research: Optimizing the use of surge capacity. Healthc Policy 3(3) (2008) 75–88.

[70] Paul J., MacDonald L., Determination of number of dedicated OR's and supporting pricing mechanisms for emergent surgeries. J Oper Res Soc 64(6) (2012) 912–24.

[71] Peck J.S., Gaehde S. a, Nightingale D.J., Gelman D.Y., Huckins D.S., Lemons M.F., Dickson E.W., Benneyan J.C., Generalizability of a simple approach for predicting hospital admission from an emergency department. Acad Emerg Med 20(11) (2013) 1156–63.

[72] Persson M.J., Persson J.A., Analysing management policies for operating room planning using simulation. Health Care Manag Sci 13(2) (2010) 182–91.

[73] Pham D.-N., Klinkert A., Surgical case scheduling as a generalized job shop scheduling problem. Eur J Oper Res 185(3) (2008) 1011–25.

[74] Roberts T.T., Vanushkina M., Khasnavis S., Snyder J., Papaliodis D.N., Rosenbaum A.J., Uhl R.L., Roberts J.T., Bagchi K., Dedicated orthopaedic operating rooms: Beneficial to patients and providers alike. J Orthop Trauma (in press) (2014)

[75] Ryckman F., Adler E., Anneken A., Bedinghaus C., Clayton P.J., Hays K.R., Lee B., Morillo-delerme J.W., Schoettker P.J., Yelton P.A., Kotagal U.R., Cincinnati Children's Hospital Medical Center: Redesigning perioperative flow using operations management tools to improve access and safety. In: Litvak E., ed. Managing Patient Fow in Hospitals: Strategies and Solutions. Oakbrook Terrace, IL: Joint Commission International, 2009:97–111.

[76] Van Ryzin G., Talluri K.T., An Introduction to Revenue Management. TutORials Oper Res (2005) 142–94.

[77] Samudra M., First doctoral seminar. 2013 http://feb.kuleuven.be/healthcare/seminar_1/.

[78] Sandbaek B.E., Helgheim B.I., Larsen O.I., Fasting S., Impact of changed management policies on operating room efficiency. BMC Health Serv Res 14(1) (2014) 1–10.

[79] Santibáñez P., Begen M., Atkins D., Surgical block scheduling in a system of hospitals: An application to resource and wait list management in a British Columbia health authority. Health Care Manag Sci 10(3) (2007) 269–82.

[80] Smith C.D., Spackman T., Brommer K., Stewart M.W., Vizzini M., Frye J., Rupp W.C., Re-engineering the operating room using variability methodology to improve health care value. J Am Coll Surg 216(4) (2013) 559–68.

[81] Sobolev B.G., Brown P.M., Zelt D., Fitzgerald M., Priority waiting lists: Is there a clinically ordered queue? J Eval Clin Pract 11(4) (2005) 408–10.

[82] Steins K., Persson F., Holmer M., Increasing utilization in a hospital operating department using simulation modeling. Simulation 86(8-9) (2010) 463–80.

[83] Strum D.P., May J., Vargas L.G., Modeling the uncertainty of surgical procedure times. Anesthesiology 92(4) (2000) 1160–7.

[84] Stuart K., Kozan E., Reactive scheduling model for the operating theatre. Flex Serv Manuf J 24(4) (2011) 400–21.

[85] Surgeons R.A.C. of, The case for the separation of elective and emergency surgery. 2011 http://www.surgeons.org/media/307115/sbm_2011-05-24_separating_elective_and_emergency_surgery.pdf.

[86] Tancrez J.-S., Roland B., Cordier J.-P., Riane F., How stochasticity and emergencies disrupt the surgical schedule. In: McClean S., Millard P., El-darzi E., Nugent C.D., eds. Intelligent Patient Management. Springer Berlin Heidelberg, 2009:221–39.

[87] Tancrez J.-S., Roland B., Cordier J.-P., Riane F., Assessing the impact of stochasticity for operating theater sizing. Decis Support Syst 55(2) (2013) 616–28.

[88] Testi A., Tanfani E., Valente R., Ansaldo G.L., Torre G.C., Prioritizing surgical waiting lists. J Eval Clin Pract 14(1) (2008) 59–64.

[89] Testi A., Tànfani E., Tactical and operational decisions for operating room planning: Efficiency and welfare implications. Health Care Manag Sci 12(4) (2009) 363–73.

[90] Vanberkel P.T., Blake J.T., A comprehensive simulation for wait time reduction and capacity planning applied in general surgery. Health Care Manag Sci 10(4) (2007) 373–85.

[91] Vansteenkiste N., Lamote C., Vandersmissen J., Luysmans P., Monnens P., De Voldere G., Kips J., Rademakers F.E., Reallocation of operating room capacity using the due-time model. Med Care 50(9) (2012) 779–84.

[92] Vieira G.E., Herrmann J.W., Lin E., Rescheduling manufacturing systems: A framework of strategies, policies and methods. J Sched 6 (2003) 39–62.

[93] Wullink G., Van Houdenhoven M., Hans E.W., van Oostrum J.M., van der Lans M., Kazemier G., Closing emergency operating rooms improves efficiency. J Med Syst 31(6) (2007) 543–6.

[94] Zhang B., Murali P., Dessouky M.M., Belson D., A mixed integer programming approach for allocating operating room capacity. J Oper Res Soc 60(5) (2009) 663–73.

[95] Zhou J., Dexter F., Method to assist in the scheduling of add-on surgical cases: Upper prediction bounds for surgical case durations based on the log-normal distribution. Anesthesiology 89(5) (1998) 1228–32.

[96] Zonderland M., Boucherie R., Litvak N., Vleggeert-Lankamp C., Planning and scheduling of semi-urgent surgeries. Health Care Manag Sci 13(3) (2010) 256–67.

[97] Zonderland M., Timmer J., Optimal allocation of MRI scan capacity among competing hospital departments. Eur J Oper Res 219(3) (2012) 630–7.

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be